DOCUMENT RESUME

ED 434 037 SE 062 892

AUTHOR Allen, Bradford D.; Carific, James
TITLE The Development and Validation of an Emotion Questionnaire for the Investigation of Affect during Mathematical Problem Solving.
PUB DATE 1999-10-00
NOTE 32p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Affective Behavior; Elementary Secondary Education; *Mathematics Education; Measures (Individuals); *Problem Solving; Questionnaires
IDENTIFIERS *Polya (George)

ABSTRACT

The importance of emotion in problem solving is well established. The Emotion Questionnaire, presented here, is a 38 item instrument designed to measure important aspects of emotion problem solving. The development and validation of the Emotion Questionnaire is discussed. Specific results gleaned in developing and validating this instrument are reviewed. These results support Polya's theory of emotion during mathematical problem solving. (Contains 27 references.) (Author/ASK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: SE
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. TM

ED 434 037

The Development and Validation of an Emotion Questionnaire for the Investigation of Affect During Mathematical Problem Solving

Bradford D. Allen, Florida Institute of Technology
James Carifio, UMASS at Lowell

ABSTRACT

The importance of emotion in problem solving is well established. The Emotion Questionnaire, presented here, is a 38 item instrument designed to measure important aspects of emotion during mathematical problem solving. The development and validation of the Emotion Questionnaire is discussed. Specific results gleaned in developing and validating this instrument are reviewed. These results support Polya's theory of emotion during problem solving.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. Allen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

October 1999

BEST COPY AVAILABLE

SE029095

Researchers and classroom teachers have long recognized that emotions influence mathematical problem solving (Dreger & Aiken, 1957), (McLeod, 1988, 1989). As Skemp (1971) and others have pointed out, emotion can be a help or hinder in solving problems. The effects of emotions in problem solving, therefore, need to be further investigated. The Emotion Questionnaire, presented here, is a 38 item instrument which may be used to evaluate various aspects of emotion during mathematical problem solving.

EMOTION QUESTIONNAIRE DEVELOPMENT

Gable and Wolf's (1993) approach to instrument development in the affective domain was used to develop the Emotion Questionnaire, a semantic differential which measures emotion during problem solving. The theoretical concepts that manifest and define emotion are operationalized in the Emotion Questionnaire. These concepts come from Polya's and Mandler's theories of emotion.

Polya (1968, 1981) describes specific metacognitive evaluations which occur at each step of solving a problem. According to Polya, every step is accompanied by the evaluations of relevancy, proximity, and quality. Polya claims that emotion arises from these three types of evaluations. These evaluations described by Polya are based on the questions "am I focused on the problem?", "how close to the solution am I?", and "how good is my plan and what is its likely fate?" More specifically, relevancy is the evaluation of how important the problem is to the problem solver or how important the idea or guess is in getting to a solution, proximity is the evaluation of how close or far away the

solution is, and quality is the evaluation of how good the problem solver's own performance is in finding a solution and what is the likely outcome of the guess and/or overall performance.

A fundamental assumption of Mandler's (1975, 1982, 1984, 1989) theory is that emotion arises from the combination of evaluation and physiological activity. Physiological activity is a necessary and measurable part of mobilization. Mandler claims that physiology is nonspecific in that it contributes nothing to the evaluation of the situation. Instead, physiological activity only provides a visceral or energized stimulation that determines the intensity of emotion. Unlike physiology, the process of evaluating a situation (in perticular, how an interruption is interpreted) determines only the quality or tone of emotion. Combined together, evaluation and physiology are the major factors which give rise to emotion. Emotion intensity depends on the physiology level, and the emotional tone (whether the emotion is agreeable or disagreeable) depends only on the evaluation process.

Five major components of emotion were derived from the above theories and designed into the Emotion Questionnaire. These components are physiological activity, emotion, and three forms of evaluation: relevancy, proximity, and quality. According to the literature, these concepts are separate and distinct, though in many real instances, they are not. Five subscale categories in the Emotion Questionnaire operationalize these five major concepts. In summary, relevancy, proximity, quality, physiological activity, and emotion are the five constructs measured by five subscale categories in the emotion questionnaire.

INSTRUMENT DESIGN

The subscale components of the Emotion Questionnaire were designed from samples of items from established instruments that measure those components. Carifio (1995) shows empirically that random samples of items from validated instruments have the same psychometric properties (given adjustments for sampling error) as the full length instruments.

The physiological component of the semantic differential uses frequently appearing key words from several instruments that measure physiological activity. These instruments are Cattell's (1966) *Self-Ratings of Anxiety Survey*, Mandler and Watson's (1966) measurements of obsessive or ruminative thoughts irrelevant to problem completion, Spielberger, Gorsuch, and Lushene's (1970) *State-Trait Anxiety Inventory*, and Richardson and Suinn's (1972) *Mathematics Anxiety Rating Scale*.

The emotion component of the semantic differential uses frequently appearing key words from Schachter's (1966) emotion questionnaire, Bush's (1973) study of adjectives that denote feelings, Russell's (1979) affect semantic differential, Abella and Heslin's (1989) *Emotion and Coping Scale*, and Bolin and Dodder's (1992) *Affect Balance Scale*.

The three evaluation components of the Emotion Questionnaire are derived from words distilled from Polya's (1945, 1968) discussions about evaluation during problem solving.

Using the words selected from the above emotion studies, bipolar word pairs were created using Osgood and Suci's (1969) semantic differential technique, Osgood, May, and Miron's (1975)

affective interpretation of the semantic differential, the Dale-Chall list of 3000 Familiar Words (Dale and Chall 1948), and Shaver, Schwartz, Kirson, and O'Connor's (1987) cluster analysis of 135 emotion names.

The 38 bipolar word pairs on the semantic differential were formed using key words that appeared frequently in the literature. The bipolar word pairs were grouped into six emotion category subscales. These categories operationalize the emotion and emotion related concepts and are presented in Table 1. Numbers in the position column indicate where items appear on the Emotion Questionnaire. An "R" in the reverse scoring column indicates the item is reverse scored. A "II" to the right of the word pair means the item is targeted at the process of getting the answer rather than at the feelings resulting from working on the problem. (The first thirty-two items refer to "How do you feel from working on this problem?" and the last six pairs refer to "Getting the answer to this problem is.")

Table 1. Category subscales of the semantic differential

category	position	bipolar word pair	reverse scoring
RELEVANCY	1	focused - distracted	R
	6	riveted - wandering	R
	11	bored - interested	
	16	attracted - repelled	R
	21	indifferent - eager	
	34	irrelevant - important (II)	
	37	interesting - dull(II)	R

Table 1. Category subscales of the semantic differential

category	position	bipolar word pair	reverse scoring
PROXIMITY	2	slow - fast	
	7	confused - not confused	
	12	certain - uncertain	R
	17	optimistic - pessimistic	R
	22	hopeless - hopeful	
	33	difficult - easy (II)	
	36	in sight - hidden(II)	R
	38	near - far(II)	R
QUALITY	3	unsure - confident	
	8	guessing - knowing	
	13	dumb - smart	
	18	able - unable	R
	23	correct - incorrect	R
	26	precise - imprecise	R
	29	hazy - clear	
	31	accurate - inaccurate	R
PHYSIOLOGICAL ACTIVITY	4	excited - calm	R
	9	tense - relaxed	R
	14	at ease - restless	
	19	composed - nervous	
	24	tranquil - perturbed	
	27	contented - dissatisfied	
	32	worried - peaceful	R
EMOTION	5	distressed - delighted	
	10	good - bad	R
	15	successful - unsuccessful	R
	20	irritated - soothed	
	25	frustrated - satisfied	
	28	proud - shamed	R
	30	pleasant - unpleasant	R
	35	annoying - pleasing (II)	

II - The sentence being rated is "getting the answer to this problem is."

Each word pair on the Emotion Questionnaire is separated by seven spaces as in the sample item below.

5. distressed___:___:___:___:___:___:___delighted

METHODOLOGY

Prior to answering the Emotion Questionnaires, students were instructed to rate the target sentences in the Emotion Questionnaire by placing a mark in one and only one of the spaces as follows: If students felt that one end of the scale was very closely related to the sentence being rated, they were instructed to place a mark in the space right next to the word at that end. Students were instructed to place a mark one space over from the word at the end if it was felt that one end of the scale was closely related to the sentence being rated. If it was felt that one end of the scale was slightly related to the sentence being rated, they were instructed to place a mark two spaces over from the word at that end. Students were instructed to place a mark in the middle space if they felt both sides of the scale are equally related to the sentence or if the pair of words was completely irrelevant. Students were instructed that marks should be made in the middle of the spaces and not on the boundaries, that no more than one mark should be made on any scale, and no scale should be omitted. The specific instructions to students are given in the Appendix.

Two hundred nine undergraduate students took part in an experiment which comprised of answering a math affect trait questionnaire, solving or trying to solve two math problems, and answering either two or six emotion questionnaires. One math problem was a low difficulty problem and the other was a medium difficulty problem. The problems were presented in random order.

While solving the problems, 189 students completed six Emotion Questionnaires. A questionnaire was filled out after each problem was first read, half way completed, and completed or given up on. The remaining twenty students filled out one Emotion Questionnaire after completing or giving up on each problem. It was made clear to all students that all written responses should be restricted to thoughts, emotions, and feelings that result from their problem solving only.

In all, 206 complete Math Affect Trait Questionnaires and 1084 complete Emotion Questionnaires were collected. Of the 189 subjects designated to fill out six Emotion Questionnaires, 173 answered the third questionnaire. Of these, 171 also answered a math affect trait questionnaire. Of these, 152 completed six Emotion Questionnaires, the trait questionnaire, and worked on the two problems.

The Emotion Questionnaire was a semantic differential with seven answer steps between each pair of words; thus, the questionnaire's items were scored from 1 to 7 in correspondence with the seven steps. A "1" was given for a check in the space next to the word on the left of the item and a "7" was given for a check next to the word of opposite meaning on the right. The middle space represented neutral and was scored as "4". Item numbers 1, 6, 10, 12, 14 to 19, 23, 24, 26, 27, 28, 30,31, 36, 37, and 38 on the Emotion Questionnaire were reverse scored. Higher scores on the Emotion Questionnaire implied higher levels of relevancy of the problem to the individual, closer proximity to the solution, better quality of work, more physiological activity, and positive emotion.

RESULTS

To refine the Emotion Questionnaire into a valid and reliable instrument, item analyses, factor analyses, and reliability analyses were performed on items and category subscales. Three items were eliminated because of low variance. Because there were only two principle component subscales which contributed more than five percent to total variance, the five-subscale design deriveded from theory and literature is retained. High reliability and validity coefficients were found for these five subscales.

Table 2 contains summary descriptive statistics for the Emotion Questionnaire. When items were checked for aberrant means or low standard deviations, item numbers 6, 16, and 28 were found to have low standard deviations. These items were eliminated from further analyses. A large proportion of the Emotion Questionnaire item distributions were either skewed or platykurtic. The bimodal nature of responses which was predicted (see Allen & Carifio, 1995) generally did not occur across pairs of opposite meaning words. Most items received neutral responses. This may have been due to subjects placing no importance on the math problems or questionnaire items, subjects being unable to decide, and/or subjects feeling neutral about the anchor words.

Table 2. Summary Statistics for Emotion Questionnaire items from pooled data over problems, occasions and experimental formats (N=1082).

Item#	Mean	Std Dev	Kurtosis	Skewness	Word pair
1	4.87	2.03	-.95	-.59	focused/distracted
2	3.88	2.04	-1.22	.11	slow/fast
3	4.00	2.26	-1.50	.05	unsure/confident

Table 2. Summary Statistics for Emotion Questionnaire items from pooled data over problems, occasions and experimental formats (N=1082).

Item#	Mean	Std Dev	Kurtosis	Skewness	Word pair
4	3.95	1.76	-.75	.11	excited/calm
5	3.89	1.83	-.87	.14	distressed/delighted
6	3.96	1.69	-.57	-.02	riveted/wandering
7	3.86	2.14	-1.35	.16	confused/not confused
8	3.83	2.15	-1.36	.11	guessing/knowing
9	4.42	1.96	-1.13	-.19	tense/relaxed
10	4.54	1.90	-.95	-.31	good/bad
11	4.43	1.87	-.94	-.26	bored/interested
12	4.10	2.12	-1.38	-.03	certain/uncertain
13	4.46	1.85	-.88	-.21	dumb/smart
14	4.52	1.87	-.99	-.25	at ease/restless
15	4.25	2.07	-1.26	-.16	successful/ unsuccessful
16	4.34	1.69	-.50	-.22	attracted/repelled
17	4.52	1.99	-1.08	-.34	optimistic/ pessimistic
18	4.43	2.11	-1.26	-.30	able/unable
19	4.79	1.72	-.57	-.43	composed/nervous
20	3.74	1.84	-.87	.17	irritated/soothed
21	4.06	1.84	-.93	-.04	indifferent/eager
22	4.40	1.92	-.99	-.24	hopeless/hopeful
23	4.25	2.07	-1.24	-.17	correct/incorrect
24	4.28	1.80	-.80	-.10	tranquil/perturbed
25	3.93	2.03	-1.18	.12	frustrated/satisfied
26	4.29	1.92	-1.05	-.14	precise/imprecise
27	4.21	1.96	-1.10	-.12	contented/ dissatisfied

Table 2. Summary Statistics for Emotion Questionnaire items from pooled data over problems, occasions and experimental formats (N=1082).

Item#	Mean	Std Dev	Kurtosis	Skewness	Word pair
28	4.54	1.63	-.50	-.15	proud/shamed
29	4.02	2.09	-1.33	.04	hazy/clear
30	4.41	1.82	-.89	-.17	pleasant/unpleasant
31	4.19	2.07	-1.26	-.10	accurate/inaccurate
32	4.38	1.79	-.85	-.10	worried/peaceful
33	3.67	2.14	-1.30	.25	difficult/easy
34	3.79	2.00	-1.13	.06	irrelevant/important
35	3.57	2.02	-1.07	.27	annoying/pleasing
36	4.01	2.12	-1.31	-.01	in sight/hidden
37	4.45	1.91	-.85	-.40	interesting/dull
38	4.17	2.14	-1.30	-.15	near/far

Table 3 presents the means and standard deviations of the subscale totals which were hypothesized to measure the emotion concepts. As can be seen from the table, relevancy, proximity, and quality evaluations, self-reported physiological activity, and emotion were all stable over time and problems. It makes sense that a subject's evaluation of proximity to the solution should be greatest at the end of the problem; only slight changes in proximity were reported, however. Subjects found the easy problem slightly more relevant than the difficult problem, and they deemed the quality of their work highest at the end. Emotions were most positive at the end (after 15 minutes) of working on the problems.

Table 3. Emotion Questionnaire subscale means and standard deviations at beginning, middle and end of easy and difficult problems (N=152).

problem	Time	relevancy 5 items neutral =20		proximity 8 items neutral =32		quality 8 items neutral =32		physiology 7 items neutral =28		emotion 7 items neutral =28	
		mean	std dev	mean	std dev	mean	std dev	mean	std dev	mean	std dev
easy	1	22.1	6.6	33.4	13.0	33.7	13.5	30.9	8.8	28.2	10.5
	2	22.2	7.4	33.5	14.6	33.4	15.2	30.1	9.5	28.3	12.1
	3	22.0	8.0	34.7	15.7	36.1	17.3	31.6	10.9	30.2	13.6
difficult	1	21.1	6.3	30.9	12.0	32.1	12.1	30.7	8.0	27.8	9.9
	2	20.4	6.5	30.6	12.7	31.3	13.5	29.6	8.9	26.4	10.7
	3	20.4	8.2	32.9	14.2	33.8	15.9	30.7	9.7	28.5	12.9

To better evaluate the hypothesized subscales, correlations between items and subscale total scores were checked to be sure the items on each subscale predicted the subscale total. Correlations between all subscales were examined. Subscale scores for each questionnaire were factor analyzed separately and together. The distributions of subscale scores were checked. These procedures were done in terms of time, problem difficulty, and problem correctness.

Table 4. Emotion Questionnaire subscale item means (neutral=4) and standard deviations at beginning, middle and end of easy and difficult problems
N(time=1,2)=152, N(time=3)=173.

problem	Time	relevancy 5 items		proximity 8 items		quality 8 items		physiology 7 items		emotion 7 items	
		mean	std dev	mean	std dev	mean	std dev	mean	std dev	mean	std dev
easy	1	3.5	.8	4.6	1.4	4.6	1.5	4.5	1.2	3.9	1.2
	2	3.5	.9	4.8	1.6	4.8	1.7	4.6	1.2	4.0	1.4
	3	3.4	.9	4.8	1.7	4.9	1.8	4.6	1.4	4.2	1.5
difficult	1	3.3	.8	4.2	1.4	4.3	1.4	4.5	1.1	3.7	1.2
	2	3.2	.9	4.1	1.4	4.2	1.4	4.3	1.2	3.5	1.2
	3	3.3	.9	4.4	1.5	4.6	1.6	4.5	1.2	4.0	1.4
tot		3.4	.9	4.5	1.5	4.6	1.6	4.5	1.2	3.9	1.3

Table 5. Correlations between Emotion Questionnaire subscale totals at the last observation on the easy problem (N=152).

	relevancy	proximity	quality	physiology	emotion
relevancy	--	.69	.67	.56	.72
proximity		--	.96	.82	.94
quality			--	.84	.95
physiology				--	.87
emotion					--

Table 6. Correlations between Emotion Questionnaire subscale totals at the last observation on the difficult problem (N=152).

	relevancy	proximity	quality	physiology	emotion
relevancy	--	.65	.70	.53	.74
proximity		--	.94	.75	.90
quality			--	.79	.94
physiology				--	.82
emotion					--

Tables 5 and 6 presents correlations between the Emotion Questionnaire subscale totals at the last observation on the easy and difficult problems. High correlations were found between all subscales, and in particular, between proximity, quality, and emotion. The high correlations between physiology and the other subscales seem to indicate that either the physiology scale is not measuring pure non-cognitive physiological response or there are intervening variables that were not identified.

Table 7. Kurtosis and skewness coefficients of Emotion Questionnaire subscales at the beginning and end of the easy and difficult problems (N=152).

problem	Time	relevancy 5 items	proximity 8 items	quality 8 items	physiology 7 items	emotion 7 items
easy	1	kt= -.5 sk= -.2	kt= -.7 sk= -.2	kt= -.8 sk= -.1	kt= -.3 sk= -.1	kt= -.6 sk= -.1
	2	kt= -.5 sk= -.2	kt=-1.1 sk= 0	kt=-1.2 sk= 0	kt= -.4 sk= 0	kt=-1.0 sk= +.1
	3	kt= -.8 sk= -.1	kt=-1.3 sk= -.2	kt=-1.4 sk= -.3	kt= -.9 sk= -.1	kt=-1.2 sk= -.2
difficult	1	kt= +.1 sk= -.1	kt= -.7 sk= +.2	kt= -.7 sk= -.1	kt= -.5 sk= -.1	kt= -.3 sk= +.1
	2	kt= -.3 sk= 0	kt= -.9 sk= +.2	kt= -.8 sk= +.3	kt= -.7 sk= +.1	kt= -.6 sk= +.2
	3	kt =-.9 sk= +.1	kt=-1.2 sk= 0	kt= -1.3 sk= -.1	kt= -.7 sk =-.3	kt=-1.2 sk= -.1

kt=kurtosis, sk=skewness

Table 7 shows kurtosis and skewness coefficients at the beginning and end of both problems. The coefficients show a tendency for Emotion Questionnaire subscales to be asymmetric and flatter than normal distribution. Thus, much of the statistics such as correlations and factor analyses were done with the assumption of normally distributed variables not fully met. This, combined with the highly correlated observations and truncated variance increased the chance of type I errors (rejecting H_0 when true). Therefore, the critical regions on statistical tests needed to be very small (for example $p<.0005$).

The correlations between Emotion Questionnaire subscale items and their corresponding subscale totals were high. However, because each subscale item contributes to its subscale total, an

autocorrelation effect would have produced larger correlations than would otherwise be expected. Item total correlations for relevancy ranged from .67 to .80 with a mean of .75. Item total correlations for proximity ranged from .69 to .86 with a mean of .80. Item total correlations for quality ranged from .81 to .92 with a mean of .85. Item total correlations for physiology ranged from .21 to .84 with a mean of .71. Item number 4 lowered the average item correlation for physiology. Item total correlations for emotion ranged from .79 to .90 with a mean of .84.

Table 8 presents a factor analysis of the Emotion Questionnaire subscales at the conclusion (after 15 minutes of work) of the difficult problem. Factor analyses at all other problem and time combinations gave similar results. Principle components factoring with varimax orthogonal rotation was used. Subscale totals were Kaiser normalized and 1's were used on the matrix diagonal. The eigenvalue cutoff was set at one. The cumulative percent variance accounted for by the single retained factor was 81%.

Table 8. Principle components factor analysis with varimax orthogonal rotation of Emotion Questionnaire subscales at conclusion of difficult problem (N=189).

subscale	factor I	communality
emotion	.97	.58
quality	.96	.88
proximity	.94	.93
physiological	.86	.73
relevancy	.76	.94
percent variance	81%	81%

The factor loadings show that variation in emotion is mirrored by the other scales. However, emotion's low communality relative to the other scales indicates that emotion has less in common with those scales and thus, is poorly predicted by the other scales.

Comparing a priori item categories with empirically derived constructs identified by factor analysis provides a means of examining the construct validity of a questionnaire. Before factor analyzing all the Emotion Questionnaire items, it was questionable as to whether the "How do you feel from working on this problem?" items in the Emotion Questionnaire (items 1 through 32) should be factored separately from the "Getting the answer to this problem is" (items 33 through 38). To resolve this, the two sets of items were factored and non-orthogonally rotated separately and together. When items 33 through 38 were factor analyzed separately, two main factors accounting for 76% of the variance were identified. The two factors were made up of proximity evaluation items (items 33,38,36 and 35) and relevancy evaluation items (items 34 and 37). When items 33 through 38 were factor analyzed with the other Emotion Questionnaire items, the proximity evaluation items clustered with the other proximity evaluation items and the relevancy evaluation items clustered with the other relevancy evaluation items. Furthermore, the correlation between these two factors were virtually the same in both groups. It was therefore concluded that the proximity and relevancy evaluation items 33 through 38 were measuring the same types of evaluations as the proximity and relevancy evaluation items 1 through 33 so could be factor analyzed together.

When the Emotion Questionnaire items were factor analyzed, the items grouped together along one dominant factor and one very weak factor. The variance of the weak factor increased from 4 percent at the beginning of the problems to 9 percent at the end of the problems when emotion responses were at extremes. The responses for both problems were pooled together across measurement times to give a sufficiently large sample size of N=378 with which to factor thirty-five items. When the items were factored separately for each problem using sample sizes of N=189, the resulting factor structures were similar to the pooled analysis. The three two-factor structures each accounted for roughly 63 percent of the questionnaire variance with the dominant and weak factors accounting for 53 percent and 9 percent respectively. When the items were sorted according to their loadings on the dominant factor and the item order was compared, the correlations were fairly high. Table 9 presents the Spearman rank order correlations for the highest loading eighteen items on the dominant factor for individual and pooled problems.

Table 9. Spearman rank order correlations of the eighteen highest loading items on the dominant factor for three factor analyses using individual and pooled problems (N=18).

	problem 1	problem 2	problems 1&2
problem 1	--	.70	.67
problem 2		--	.77
problems 1&2			--

The above correlations indicate that when the two problems are pooled together, the items are weighted on the main factor in a way similar to the weighting for individual problem factor

analyses. Furthermore, some of the difference in the factor item order may be due to the small N when just individual problems are used. With minimal effect on factor analysis results due to pooling, the factor structure presented below is computed from pooled problem data collected at problem completion.

Table 10 lists the items of each subscale hypothesized to measure the above concepts and embedded in the Emotion Questionnaire. Alpha internal consistency and average communality (h^2) estimates of reliability of the hypothesized subscales are also given. The table also presents empirically derived factor loadings on the two main factors. The loadings were derived from Emotion Questionnaire data collected at problem completion when reported emotions were at extremes. Data was pooled across problems giving a sample size of N=378. Profile analyses on the two problems highlighted the differences between the two problems.

The factor loadings were derived using principle components factoring with varimax orthogonal rotation. Items were Kaiser normalized and 1's were used on the matrix diagonal. In order to combine all factors that contribute less than five percent to the total variance, the eigenvalue cutoff was set at 1.4. Only cases with non-missing values on all items were analyzed. Loadings less than .5 are not shown. The cumulative percent variance accounted for by the two retained factors was 60% with 53% and 7% attributed to factors I and II respectively.

Table 10. Hypothesized Subscales in the Emotion Questionnaire with Internal Consistency Estimates, Unsorted Orthogonally Rotated Factor Loadings, Item Communalities (N=378).

Subscale items	factor I	factor II	h2
RELEVANCY			
1 focused-distracted (R)	.54		.36
11 bored-interested		.56	.33
21 indifferent-eager		.58	.37
34 irrelevant-important		.56	.30
37 interesting-dull (R)			.22
		α=.84	h^2=.32
PROXIMITY			
2 slow-fast		.67	.57
7 confused-not confused		.65	.57
12 certain-uncertain (R)	.80		.78
17 optimistic-pessimistic(R)	.75		.67
22 hopeless-hopeful		.69	.65
33 difficult-easy		.68	.64
36 in sight-hidden(R)	.66		.55
38 near-far(R)	.72		.64
		α=.93	h^2=.63
QUALITY			
3 unsure-confident		.71	.70
8 guessing-knowing		.73	.71
13 dumb-smart		.71	.68
18 able-unable (R)	.80		.78
23 correct-incorrect (R)	.81		.79
26 precise-imprecise (R)	.80		.75
29 hazy-clear		.73	.72
31 accurate-inaccurate (R)	.80		.79
		α=.96	h^2=.74
PHYSIOLOGICAL			
4 excited-calm (R)			.00
9 tense-relaxed (R)		.63	.50
14 at ease-restless	.72		.59
19 composed-nervous	.66		.48
24 tranquil-perturbed	.67		.52
27 contented-dissatisfied	.79		.77
32 worried-peaceful (R)		.60	.47

Table 10. Hypothesized Subscales in the Emotion Questionnaire with Internal Consistency Estimates, Unsorted Orthogonally Rotated Factor Loadings, Item Communalities (N=378).

Subscale items	factor I	factor II	h2
		α=.84	h^2=.48
EMOTION			
5 distressed-delighted		.75	.70
10 good-bad (R)	.80		.76
15 successful-unsuccessful (R)	.81	.37	.81
20 irritated-soothed		.71	.64
25 frustrated-satisfied		.74	.74
30 pleasant-unpleasant (R)	.77		.71
35 annoying-pleasing		.75	.68
		α=.84	h^2=.72
percent variance	53%	7%	60%

A factor analysis with an orthogonal factor rotation was used to find independent constructs measured by the questionnaire. This analysis found the Emotion Questionnaire consists of one primary factor and a weak second factor. The primary factor accounted for 53% of the total variance. The second factor accounted for 7% of the variance. These empirically derived factors revealed that subjects failed to make independent discriminations on items and basically had a single global response to the questionnaire. However, the five subscales in the Emotion Questionnaire were not required to be independent. It was decided, therefore, not to use a factor model of this instrument and to use the hypothesized subscales.

VALIDATION OF POLYA'S THEORY OF EMOTION

Polya claims that every step in problem solving is accompanied by the evaluations of relevancy, proximity, and quality, and that emotion arises from these evaluations. Polya

also claims that more sophisticated people experience more differentiated emotion during problem solving.

MANOVA and ANOVA tests showed that problem correctness, a surrogate variable for sophistication, had a significant effects on relevancy, proximity, quality, physiology and emotion as a group (Mult.F=5.42, hyp.df=10, err.df=364, p<.0005, N=189). An examination of individual item means and standard deviations for Emotion Questionnaire subscales for the difficult problem at the third observation time revealed that high scoring subjects had relevancy, proximity, quality, physiology and emotion subscale scores at least one standard deviation above their low scoring counterparts. These findings support Polya's claim that people who are more sophisticated in the sense of being better problem solvers have more differentiated feelings during problem solving.

An affect trait profile was defined as a contrast between anxiety and the sum of math interest and self-esteem. Because this trait profile was defined in terms of variables related to knowing, learning, testing, and education, it was used as a surrogate variable for sophistication. It was observed that increases in trait profile values resulted in significant increases in Polya's evaluation variables as well as in emotion (F=23, df=10,151, p<.001, N=50). This supports Polya's claim that people who are sophisticated in the sense of having low-anxiety, high-math-interest, and high-self-esteem have more differentiated emotions. Furthermore, the relatively small increases in the physiology variable support Polya's decision not to include a physiology factor in his emotion model.

Polya's claim that correlations between relevancy, proximity, and quality should be high was supported ($.67<r<.95$, $N=152$). High correlations were found between the evaluation variables and emotion ($.73<r<.94$, $N=152$), and in particular, between proximity, quality, and emotion ($r=.92$, $r=.94$ resp., $N=152$). These results are consistent with Polya's claim that his three forms of evaluation lead to emotion.

Polya's claim that relevancy, proximity, and quality evaluations lead to emotion was further supported by regression analyses which showed that relevancy, proximity, quality, and perceived physiology explained roughly 93 percent of the variance in the emotion variable. Coefficients in these regression equations indicated that relevancy and quality evaluations may be roughly twice as important in Polya's theory as is proximity.

Repeated measures ANOVA and MANOVA tests supported Polya's claim that relevancy, proximity, and quality are ongoing evaluations that lead to emotion. Significant differences were found in Polya's evaluation variables and in emotion over time of measurement (Mult.F=3.28, hyp.df=10, err.df=142, $p<.001$, $N=152$) and problem type (Mult.F=3.47, hyp.df=5, err.df=147, $p<.005$, $N=152$). These results indicate that evaluations and emotion are dynamic variables that change with time and problem type.

Polya's decision to include only three types of evaluations during problem solving was confirmed by a factor analysis of evaluation items. The derived factor structure showed three independent evaluation categories that measured quality, relevancy, and proximity. Though Polya did not discuss the relative importance of his three evaluations, the factor analysis

revealed that the three factors were not of equal importance in terms of variance. Quality-of-work contributed almost seven times as much (55%) to the total variance as problem-relevancy contributed (8%) and over nine times as much as proximity-to-the-solution (6%) contributed.

A non-orthogonal factoring of all Emotion Questionnaire items showed quality and proximity aligned with emotion. Perceived physiology items appeared as a separate factor accounting for only four percent of the total variance in the questionnaire. This supports Polya's decision not to include physiology as a component of emotion during problem solving.

SUMMARY

Two-hundred-nine undergraduate students took part in this experiment. Each student filled out a math affect trait questionnaire and two or more Emotion Questionnaires while solving two math problems. After examining the response distributions, three items from the Emotion Questionnaire were eliminated from further analysis. This improved the questionnaire subscale correlations and factor analysis results. A poor choice of items on a subscale could easily cause low variance on that scale.

The Emotion Questionnaire items factored into one dominant factor and one weak factor. Nine items had factor loadings greater than .75 on the main factor. These items were from Polya's three evaluations and from emotion. All of these items keyed on the target statement "How do you feel from working on this problem?" The high loading items were #23 (correct-incorrect), #15 (successful-unsuccessful), #12 (certain-uncertain), #31

(accurate-inaccurate), #18 (able-unable), #10 (good-bad), #26 (precise-imprecise), #27 (contented-dissatisfied), and #30 (pleasant- unpleasant). The fact that subjects did not differentiate between these items made the factor analysis structure unusable. Furthermore, after experimenting with the factor analysis procedure by reducing the number of items on the main factor, it was evident that the factor structure with one dominant and one weak factor which accounted for 54% and 9% of the total variance did not change significantly until the number of items on the main factor was reduced to four or fewer items. Therefore, the original subscale design which was hypothesized from theory and the literature was retained for subsequent analyses. The hypothesized subscales, with their high alpha internal consistency values and estimates of reliability, were the better of the two constructions both empirically and logically.

The Emotion Questionnaire was used to validate Polya's claims that every step in problem solving is accompanied by relevancy, proximity, and quality evaluations; that emotion arises from these evaluations; and that more sophisticated people experience more differentiated emotion during problem solving.

APPENDIX

INSTRUCTIONS TO STUDENTS

Dear Students

The purpose of this study is to measure some of your thoughts and feelings while you solve math problems. Your thoughts and feelings will be measured with two questionnaires: the Preliminary Questionnaire and the Problem Solving Questionnaire. All of your answers will be kept strictly confidential and the results will only be reported as averages of answers of all participants.

The Preliminary Questionnaire consists of twenty-five statements with six choices each which range from "agree strongly" to "disagree strongly." Place a mark next to the choice that is closest to how you feel about the statement.

The Problem Solving Questionnaire will be filled out either twice, if you are in the procedure I group, or six times while you solve two math problems. This questionnaire consists of thirty-eight pairs of words of opposite meaning. By placing a mark between the pairs, you will rate one of two sentences. The two sentences are "How do you feel from working on this problem?" and "Getting the answer to this problem is." For each pair of words, if you feel that one end of the scale is VERY CLOSELY RELATED to the sentence being rated, place a mark in the space right next to the word at that end. If you feel that one end of the scale is QUITE CLOSELY RELATED to the sentence being rated, place a mark one space over from that end. If one end of the scale is ONLY SLIGHTLY RELATED to the sentence being rated, place a mark two spaces over from that end. Place a mark in the middle of the scale if both sides are EQUALLY RELATED to the sentence, or if the pair of words is completely IRRELEVENT to the sentence.

Place your marks in the middle of the spaces, not on the boundaries. Do not put more than one mark on any scale, and do not

omit any scale. Your responses should be ONLY your thoughts and feelings about solving the problem. Any feelings from having to stop and answer the questionnaire should not be included in your responses.

Participation in this study will involve one of three possible procedures. All participants will fill out the Preliminary Questionnaire and solve two math problems. One of those problems has two parts and if you finish the first part, go on to the next part. However, do not start the second problem until you are instructed to do so.

Procedure I participants will work on each problem for twelve minutes and then spend about three minutes filling out a Problem Solving Questionnaire. Procedure II participants will spend about seventeen minutes working on each problem but during that time will fill out three Problem Solving Questionnaires. They will fill out a Questionnaire after each problem is first read, after seven minutes, and immediately after the problem is solved or after fifteen minutes whichever comes first. After the Questionnaires are completed, keep working on the problem but do not go on to the second problem until instructed to do so. Procedure III is the same as procedure II except procedure III participants will have their heart-rates measured just prior to filling out each Questionnaire. Heart-rate measurements will be taken with a small, harmless device worn on one finger.

You will be told when it is time to fill out each Questionnaire. If you decide to stop working on a problem (whether or not you have found an answer), you should fill out a Questionnaire at that time. If you need to leave the room for a short period of time, the procedure will stop until you return. After everyone is finished, I will briefly go over the study and answer any questions.

Thank you for assisting in this study. Your help is greatly appreciated.

EMOTION QUESTIONNAIRE

How do you feel from working on this problem?

1. focused___:___:___:___:___:___:___distracted	2. slow___:___:___:___:___:___:___fast
3. unsure___:___:___:___:___:___:___confident	4. excited___:___:___:___:___:___:___calm
5. distressed___:___:___:___:___:___:___delighted	6. riveted___:___:___:___:___:___:___wandering
7. confused___:___:___:___:___:___:___not confused	8. guessing___:___:___:___:___:___:___knowing
9. tense___:___:___:___:___:___:___relaxed	10. good___:___:___:___:___:___:___bad
11. bored___:___:___:___:___:___:___interested	12. certain___:___:___:___:___:___:___uncertain
13. dumb___:___:___:___:___:___:___smart	14. at ease___:___:___:___:___:___:___restless
15. successful___:___:___:___:___:___:___unsuccessful	16. attracted___:___:___:___:___:___:___repelled
17. optimistic___:___:___:___:___:___:___pessimistic	18. able___:___:___:___:___:___:___unable
19. composed___:___:___:___:___:___:___nervous	20. irritated___:___:___:___:___:___:___soothed
21. indifferent___:___:___:___:___:___:___eager	22. hopeless___:___:___:___:___:___:___hopeful
23. correct___:___:___:___:___:___:___incorrect	24. tranquil___:___:___:___:___:___:___perturbed
25. frustrated___:___:___:___:___:___:___satisfied	26. precise___:___:___:___:___:___:___imprecise
27. contented___:___:___:___:___:___:___dissatisfied	28. proud___:___:___:___:___:___:___shamed
29. hazy___:___:___:___:___:___:___clear	30. pleasant___:___:___:___:___:___:___unpleasant
31. accurate___:___:___:___:___:___:___inaccurate	32. worried___:___:___:___:___:___:___peaceful

Getting the answer to this problem is

33. difficult___:___:___:___:___:___:___easy	34. irrelevant___:___:___:___:___:___:___important
35. annoying___:___:___:___:___:___:___pleasing	36. in sight___:___:___:___:___:___:___hidden
37. interesting___:___:___:___:___:___:___dull	38. near___:___:___:___:___:___:___far

BEST COPY AVAILABLE

REFERENCES

Abella, R. & Heslin, R. (1989). Appraisal Processes, coping and the regulation of stress-related emotion in a college examination. Basic and Applied Social Psychology, 10(4), 311-327.

Bolin, B. L. & Dodder, R. A. (1992). The affect balance scale: A reassessment among college students. Education Research Quarterly, 15(2),5-10

Bush, L. E. (1973). Individual differences multidimensional scaling of adjectives denoting feelings. Journal of Personality and Social Psychology, 25, 50-57.

Carifio, J. (1995). Sensitive data and students' tendencies to give socially desirable responses. University of Massachusetts Lowell.

Cattell, R. B. (1966). Anxiety & motivation: Theory and experiments. In C. D. Spielberger (Ed.), Anxiety and Behavior. New York: Academic Press.

Dale, E. & Chall, J. (1948). A formula for predicting reliability. Educational Research Bulletin, 27, 11-20.

Dreger, R. M. and Aiken, L. R. (1957). The identification of number anxiety in a college population. Journal of Educational Psychology, 10, 344-351.

Gable, R. K. & Wolf, M. B. (1993). Instrument development in the affective domain (2nd Ed.). Boston: Kluwen Academic Publishers.

Mandler, G. and Watson, D. L. (1966). Anxiety and the interruption of behavior. In C. D. Spielberger (Ed.), Anxiety and behavior. New York: Academic Press.

Mandler, G. (1975). Mind and emotion. New York: John Wiley.

Mandler, G. (1982). The structure of value: Accounting for taste. In M. S. Clark & S. T. Fiske (Eds.) Affect and cognition: The seventeenth annual Carnegie symposium on cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Mandler, G. (1984). Mind and body. New York: W. W. Norton.

Mandler, G. (1989). Affect and learning: Causes and consequences of emotional interaction. In D. B. McLeod & V. M. Adams (Eds), Affect and mathematical problem solving: A new perspective. NY: Springer-Verlag.

McLeod, D. (1988). Affective issues in mathematical problem solving: Some theoretical considerations. Journal for Research in Mathematics Education, 19(2), 134-141.

McLeod, D. B. (1989). Research on affect on mathematics education: A reconceptualization. National council of teachers of mathematics research report, Reston VA.

Meier, S. L. (1992). Evaluating problem solving processes. The Mathematics Teacher: 85(8), 664-674.

Osgood, C. E., May, W. H. & Miron, M. S. (1975). Cross-cultural universals of affective meaning. Urbana, Illinois: University of Illinois Press.

Osgood, C. E. & Suci, G. J. (1969). Factor analysis meaning. In J. G. Snider & C. E. Osgood (Eds.) Semantic differential technique: A source book. Chicago: Aldine Publishing.

Polya, G. (1945). How to solve it: A new aspect of mathematical method. New Jersey: Princeton University Press.

Polya, G. (1968). Mathematical discovery. New York: Wiley.

Polya, G. (1981). Mathematical discovery. (Combined paperback edition). New York: Wiley.

Richardson, F. C. & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. Journal of Counseling Psychology, 19, 551-554.

Russell, J. A. (1979). Affect space is bipolar. Journal of Personality and Social Psychology, 37(3), 345-356.

Schachter, S. (1966). The interaction of cognitive and psychological determinants of emotional state. In L. Berkowitz (Ed.), Advances in experimental psychology. New York: Academic Press.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge. Journal of Personality and Social Psychology. Washington, DC: American Psychological Association.

Skemp, R. R. (1987). The psychology of learning mathematics. Hillsdale, NJ: Lawrence Erlbaum.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologist Press.